

The Non-Regular CEO Problem

Aditya Vempaty and Lav R. Varshney

Abstract

We consider the CEO problem for non-regular source distributions (such as uniform or truncated Gaussian). A group of agents observe independently corrupted versions of data and transmit coded versions over rate-limited links to a CEO. The CEO then estimates the underlying data based on the received coded observations. Agents are not allowed to convene before transmitting their observations. This formulation is motivated by the practical problem of a firm's CEO estimating (non-regular) beliefs about a sequence of events, before acting on them. Agents' observations are modeled as jointly distributed with the underlying data through a given conditional probability density function. We study the asymptotic behavior of the minimum achievable mean squared error distortion at the CEO in the limit when the number of agents L and the sum rate R tend to infinity. We establish a $1/R^2$ convergence of the distortion, an intermediate regime of performance between the exponential behavior in discrete CEO problems [Berger, Zhang, and Viswanathan (1996)], and the $1/R$ behavior in Gaussian CEO problems [Viswanathan and Berger (1997)]. Achievability is proved by a layered architecture with scalar quantization, distributed entropy coding, and midrange estimation. The converse is proved using the Bayesian Chazan-Zakai-Ziv bound.

Index Terms

multiterminal source coding, mean-square error, midrange estimator, Chazan-Zakai-Ziv bound

I. INTRODUCTION

Consider the problem faced by the chief executive officer (CEO) of a firm with a large portfolio of projects that each have an underlying probability of success, say drawn from a uniform distribution on $[0, 1]$. Each of her subordinates will have noisy beliefs about the risks facing the projects: random variables jointly distributed with success probability, e.g. according to a copula model (common in mathematical finance to model beliefs about risks [1]–[3]). Subordinates must convey these risks, but are not allowed to convene. The CEO has cognitive constraints that limit the information rate she can receive from her subordinates, requiring subordinates to partition risks into quantal grades like A, B, C, and D, before conveyance. Such quantized grading is typical in businesses with complex information technology projects [4]. Upon receiving information from the subordinate agents, the CEO estimates the underlying success probability to minimize mean squared error (Brier score [5]) before taking action.

This is, of course, a version of the classical problem in multiterminal source coding called the CEO problem, but for a previously unanalyzed source. The CEO problem was first introduced by Berger, Zhang, and Viswanathan [6], where they considered the source of interest to be a discrete data sequence and studied the asymptotic behavior of the minimal error probability in the limit as the number of agents and the sum rate tend to infinity. It was later extended to the case when the source sequence of interest is continuous Gaussian distributed and a quadratic distortion measure is used as a performance measure [7]. Oohama studied the sum rate distortion function of the quadratic Gaussian CEO problem and determined the complete solution to the problem [8]; the full rate-distortion region was then found independently by Oohama [9] and Prabhakaran, et al. [10]. Several extensions to this problem have been studied [11]–[17], however most of these extensions continue to deal with the quadratic Gaussian setting. Viswanath formulated a similar multiterminal Gaussian source coding problem and characterized the sum rate distortion function for a class of quadratic distortion metrics [18]. In [12], [13], the authors consider the vector Gaussian case and study the sum rate for vector Gaussian CEO problem. The related problem of determining the rate region of the quadratic Gaussian two-encoder source coding problem was solved by Wagner, Tavildar, and Viswanath [14]. Chen, et al. determined bounds on the rate region for the CEO problem with general

A. Vempaty is with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244 USA. (e-mail: avempaty@syr.edu).

L. R. Varshney is with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA. (e-mail: varshney@illinois.edu).

source distributions [15]. Eswaran and Gastpar considered the CEO problem where the source is non-Gaussian but observations are still made through additive white Gaussian noise (AWGN) channel [16].

The motivating example of belief sharing in organizations is also closely connected to studies of communicating probability values [19]–[23]. Contrary to typical work in distributed source coding for inference [24] that is concerned with compressing agents' measurements, [21] studied optimal quantization of prior probabilities for Bayesian hypothesis testing. This was extended to the case of collaborative decision making in parallel fusion settings [22], very much like the CEO problem herein but in non-asymptotic regimes. Such a problem arises in several statistical signal processing, economics, and political science settings such as human affairs, where juries or committees need to possess a common preference for two alternatives.

In sensor network settings where sensors see a phenomenon through Gaussian noise but produce censored data due to hardware limitations of the measuring device, observations might follow truncated Gaussian with bounded support. Censored sensor data renders celebrated information-theoretic results for Gaussian observations invalid.

Motivated by such applications, we consider the CEO problem with non-regular source distributions (in the sense of Bayesian estimation theory [25, p. 72]), and observations governed by a given conditional probability density function, for example, through copula models. More precisely we consider an i.i.d. source sequence $X(t)$, which follows a probability density function $f_X(x)$ with finite support \mathcal{X} , such that

$$\frac{\partial f_X(x)}{\partial x} \text{ or } \frac{\partial^2 f_X(x)}{\partial x^2}$$

either does not exist or is not absolutely integrable. We determine the asymptotic behavior of quadratic distortion as a function of sum rate in the limit of large numbers of agents and rate.

As commented by Viswanathan and Berger [7], results for discrete and continuous alphabets are not very different in most problems of information theory. However, for the CEO problem, the average distortion decays at an exponential rate for the discrete case and decays as $1/R$ for the Gaussian case, where R is the sum rate. We derive an intermediate $1/R^2$ decay rate behavior when the regularity conditions required for the Bayesian Cramér-Rao lower bound used in [7] do not hold. That is, we study the behavior of

$$\beta \triangleq \lim_{L, R \rightarrow \infty} R^2 D(R, L)$$

where L is the total number of agents, R is the sum rate, and $D(R, L)$ is the minimum achievable quadratic distortion for a fixed number of total agents L and sum rate R . Achievability is proved through a layered scheme that follows quantization, entropy coding, and midrange estimation at the CEO. The converse is proved by lower-bounding the distortion using the extended Bayesian Chazan-Zakai-Ziv bound [26], [27]. This result holds for the uniform distribution as a special case, implying that the CEO problem for sharing beliefs or probability values, attains $1/R^2$ convergence.

The remainder of the paper is organized as follows: Sec. II provides the mathematical problem formulation and the main result of the paper. Sec. III proves the direct part of the coding theorem and Sec. IV proves the converse part, using a version of the extended Chazan-Zakai-Ziv bound. Sec. V present some discussion on extensions and implications of results.

II. PROBLEM DESCRIPTION AND MAIN RESULT

Consider an i.i.d. source sequence of interest $\{X(t)\}_{t=1}^{\infty}$ drawn from a non-regular probability density function $f_X(x)$ with finite support \mathcal{X} . Without loss of generality, let the source be supported on $[0, 1]$.¹ Several agents (L) make imperfect conditionally independent assessments of $\{X(t)\}_{t=1}^{\infty}$, to obtain noisy versions $\{Y_i(t)\}_{t=1}^{\infty}$ for $i = 1, \dots, L$. The relationship between $X(t)$ and $Y_i(t)$ is governed by a conditionally independent probability density function $W_{\alpha}(y_i|x)$ for all agents, where α is the coupling parameter. This coupling parameter represents the strength of the dependence between the source $X(t)$ and the observations $Y_i(t)$. The agents separately compress their observations. The CEO is interested in estimating $X(t)$ such that the mean squared error (MSE) between the n -block source $X^n = [X(1), \dots, X(n)]$ and its estimate $\hat{X}^n = [\hat{X}(1), \dots, \hat{X}(n)]$ is minimized.

Let the source code \mathcal{C}_i^n of rate $R_i^n = (1/n) \log |\mathcal{C}_i^n|$ represent the coding scheme used by agent i to encode a block of length n of observed data $\{y_i(t)\}_{t=1}^{\infty}$. The CEO's estimate is given as $\hat{X}^n = \phi_L^n(C_1^n, \dots, C_L^n)$ where

¹Note that an extension to a general finite support is straightforward.

$\phi_L^n : C_1^n \times \dots \times C_L^n$ is the CEO's mapping. A specific achievability scheme for an example system, Fig. 1 shown in the sequel, illustrates the basic system structure.

We are interested in the tradeoff between the sum rate $R = \sum_{i=1}^L R_i^n$ and the MSE at the CEO, $D^n(X^n, \hat{X}^n)$, defined as:

$$D^n(X^n, \hat{X}^n) \triangleq \frac{1}{n} \sum_{t=1}^n (X(t) - \hat{X}(t))^2. \quad (1)$$

For a fixed set of codes, the MSE corresponding to the best estimator at the CEO is given by:

$$D^n(C_1^n, \dots, C_L^n) \triangleq \min_{\phi_L^n} D^n(X^n, \phi_L^n(C_1^n, \dots, C_L^n)).$$

Also, define the following quantities:

$$D^n(L, R) \triangleq \min_{\{C_i^n\}: \sum_{i=1}^L R_i^n \leq R} D^n(C_1^n, \dots, C_L^n), \quad (2)$$

$$D(L, R) \triangleq \lim_{n \rightarrow \infty} D^n(L, R), \quad (3)$$

and

$$D(R) \triangleq \lim_{L \rightarrow \infty} D(L, R). \quad (4)$$

To understand the tradeoff between sum rate and distortion, we study the following quantity:

$$\beta(\alpha) \triangleq \lim_{R \rightarrow \infty} R^2 D(R).$$

Let X be the generic random variable representing the source and Y_i represent the generic random variable representing agent i 's observation where X and Y_i are related through the conditional pdf $W_\alpha(y_i|x)$. We focus on observation channels which satisfy the following property, when placed in sequence with a forward test channel with an output auxiliary random variable U .

Property 1. For a given observation channel $W_\alpha(y|x)$ between X and Y , there exists a random variable U such that: X, Y , and U form a Markov chain, $X \rightarrow Y \rightarrow U$, and the conditional distribution of U given X , $f_{U|X}(u|x)$, has bounded support: $u \in [a(x), b(x)]$, where $(a+b)(x)$ is invertible and the inverse function $l(\cdot) := (a+b)^{-1}(\cdot)$ is Lipschitz continuous with Lipschitz constant $K > 0$, and further does not vanish at its end points: $\lim_{u \rightarrow a(x) \text{ or } b(x)} f_{U|X}(u|x) > 0$.

Let the set $\mathcal{S}(W)$ denote the set of random variables U which satisfy the above property for a given observation channel W . Explicit examples of channels satisfying this property are provided later in Sec. II-A.

We now state the main result of the paper; proofs are developed in the sequel.

Theorem 1. When conditional density W_α satisfies Property 1, the following relations hold:

$$\beta(\alpha) \leq \frac{2K^2}{\delta^2} \left(\min_{U \in \mathcal{S}(W)} I(Y; U|X) \right)^2 \quad (5)$$

and

$$\beta(\alpha) \geq \left(\min_{U: X \rightarrow Y \rightarrow U} I(Y; U|X) \right)^2 \int_{h=0}^{\infty} h \int_{\theta=0}^1 f_X(\theta) e^{-hg(\theta)} d\theta dh \quad (6)$$

where $K, \delta > 0$ are constants and

$$g(\theta) \triangleq \left\{ \frac{d}{d\Delta} - \left[\min_s \log \left(\int W_\alpha^s(y|\theta) W_\alpha^{1-s}(y|\theta + \Delta) dy \right) \right] \right\}_{\Delta=0} \quad (7)$$

is the first derivative of Chernoff information between the conditional densities W_α of the observation given $x = \theta$ and $x = \theta + \Delta$, evaluated at $\Delta = 0$. The minimums are taken over all non-trivial random variables to ensure that the conditional mutual information is non-zero.

Notice from the theorem, since $\beta(\alpha)$ is a finite constant, it implies that for a non-regular source distribution, in the limit of large sum rate, the distortion decays as $1/R^2$. This serves as an intermediate regime between the

exponential decay of the discrete case [6] and $1/R$ decay of the quadratic Gaussian case [7]. This result can be summarized as the fact that sharing beliefs (uniform) is fundamentally easier than sharing measurements (Gaussian), but sharing decisions is even easier (discrete). This shows the effect of the underlying source distribution on the asymptotic estimation performance. When the source has countably finite support set (discrete), we observe an exponential decay. On the other extreme, when the source has an unbounded support set (Gaussian), we observe a $1/R$ decay. Our result is for the source with bounded support (uniform, for example), and we get an intermediate result of $1/R^2$ decay. This suggests the intuitive observation that as the number of possibilities for the source (support) increases, it gets more difficult to communicate the values.

One can also note the similarity in structure of the lower bound of $\beta(\alpha)$ in this problem with other CEO problems [6], [7]. Most notably, in all cases, there is a minimization of conditional mutual information. Also, the bound here depends on Chernoff information which serves as a divergence metric similar to the Kullback-Leibler divergence from the discrete case [6] and as an information metric similar to Fisher information from the quadratic Gaussian case [7].

A. Examples of Observation Channels Satisfying Property 1

Property 1 may seem a little opaque, so here we give an illustrative example of a family of observation channels that satisfy it.

Proposition 1. *A sufficient condition for an observation channel to satisfy Property 1 is when its density $W_\alpha(y_i|x)$ is given by a copula conditional density function² and has discontinuity at end points.*

Proof: Let the end points of observation channel $W_\alpha(y_i|x)$ be denoted by $e_l(x)$ and $e_u(x)$. Due to its discontinuity at the end points, we have the following

$$\lim_{y_i \rightarrow e_l(x) \text{ or } e_u(x)} W_\alpha(y_i|x) > 0. \quad (8)$$

Consider the test channel given by $U_i = Y_i + N$, where N is a k -peak noise for $k \geq 2$, $f_N(n) = \sum_{l=1}^k p_l \delta(n - n_l)$, where $\delta(\cdot)$ is the Dirac delta function, $\sum_{l=1}^k p_l = 1$, and $n_1 < n_2 < \dots < n_k$. Then the conditional distribution of U given X , $f_{U|X}(u_i|x) = \sum_{l=1}^k p_l W_\alpha(u_i - n_l|x)$, has bounded support: $[e_l(x) + n_1, e_u(x) + n_k]$ and the values of $f_{U|X}(u_i|x)$ at the end points are given by

$$\lim_{u_i \rightarrow e_l(x) + n_1} f_{U|X}(u_i|x) \geq p_1 \lim_{y_i \rightarrow e_l(x)} W_\alpha(y_i|x) > 0 \quad (9)$$

and

$$\lim_{u_i \rightarrow e_u(x) + n_k} f_{U|X}(u_i|x) \geq p_k \lim_{y_i \rightarrow e_u(x)} W_\alpha(y_i|x) > 0. \quad (10)$$

This proves the proposition. ■

We now provide a specific example from the above family of observation channels and explicitly show that it satisfies Property 1.

Example 1. *As a specific example, consider the case when the source $X(t)$ and the observations $Y_i(t)$ are marginally distributed with uniform distribution in $(0, 1)$ and Clayton copula model is used to model the channel between the source and the observations. The conditional distribution of Y_i given X is given by the following (for $1/2 < \alpha < 1$)*

$$W_\alpha(y_i|x) = \begin{cases} (1 - \alpha)(xy_i)^{\alpha-1} (x^\alpha + y_i^\alpha - 1)^{1/\alpha-2}, & \text{for } (1 - x^\alpha)^{1/\alpha} \leq y_i \leq 1 \\ 0, & \text{otherwise.} \end{cases} \quad (11)$$

Using the test channel defined as $U_i = Y_i + N$, where N is a k -peak noise for $k \geq 2$, $f_N(n) = \sum_{l=1}^k p_l \delta(n - n_l)$, where $\delta(\cdot)$ is the Dirac delta function and $\sum_{l=1}^k p_l = 1$, would result in a density $f_{U|X}(u_i|x)$ given by (without

²A copula is a multivariate probability distribution for which the marginal probability distribution of each variable is uniform [1].

loss of generality, assume $n_1 < n_2 < \dots < n_k$:

$$f_{U|X}(u_i|x) = \begin{cases} p_1(1-\alpha)(x(u_i - n_1))^{\alpha-1} (x^\alpha + (u_i - n_1)^\alpha - 1)^{1/\alpha-2}, & \text{for } (1-x^\alpha)^{1/\alpha} + n_1 \leq u_i \leq 1+n_1 \\ p_2(1-\alpha)(x(u_i - n_2))^{\alpha-1} (x^\alpha + (u_i - n_2)^\alpha - 1)^{1/\alpha-2}, & \text{for } (1-x^\alpha)^{1/\alpha} + n_2 \leq u_i \leq 1+n_2 \\ \vdots \\ p_k(1-\alpha)(x(u_i - n_k))^{\alpha-1} (x^\alpha + (u_i - n_k)^\alpha - 1)^{1/\alpha-2}, & \text{for } (1-x^\alpha)^{1/\alpha} + n_k \leq u_i \leq 1+n_k \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

when $1+n_l < (1-x^\alpha)^{1/\alpha} + n_{l+1}$ for $l = 1, \dots, k-1$ to ensure the shifted versions of $W_\alpha(y_i|x)$ do not overlap.³

We now show that $f_{U|X}(u_i|x)$ given by (12) satisfies Property 1, which basically consists of two conditions on $f_{U|X}(u|x)$: bounded support and non-vanishing end points. $f_{U|X}(u_i|x)$ given in (12) has bounded support as $(1-x^\alpha)^{1/\alpha} + n_1 \leq u_i \leq 1+n_k$ (irrespective of whether the shifted versions overlap or not). Also, its values at the end points are given by:

$$\lim_{u_i \rightarrow (1-x^\alpha)^{1/\alpha} + n_1} f_{U|X}(u_i|x) = p_1(1-\alpha)x^{\alpha-1}(1-x^\alpha)^{1-1/\alpha}(0)^{1/\alpha-2} \rightarrow \infty > 0 \quad (13)$$

as $\alpha > 1/2$ and

$$\lim_{u_i \rightarrow 1+n_k} f_{U|X}(u_i|x) = p_k(1-\alpha)x^{-\alpha} > 0 \quad (14)$$

Hence, it satisfies Property 1.

Note the similarity between the form of this test channel and the random quantizer used by Fix for achieving the rate-distortion function of a uniform source [28].

As can be seen later in Sec. III, the achievability involves use of a test channel, Slepian-Wolf encoding and decoding, and midrange estimation at the CEO. Fig. 1 provides a block diagram outlining these steps for the example considered here.

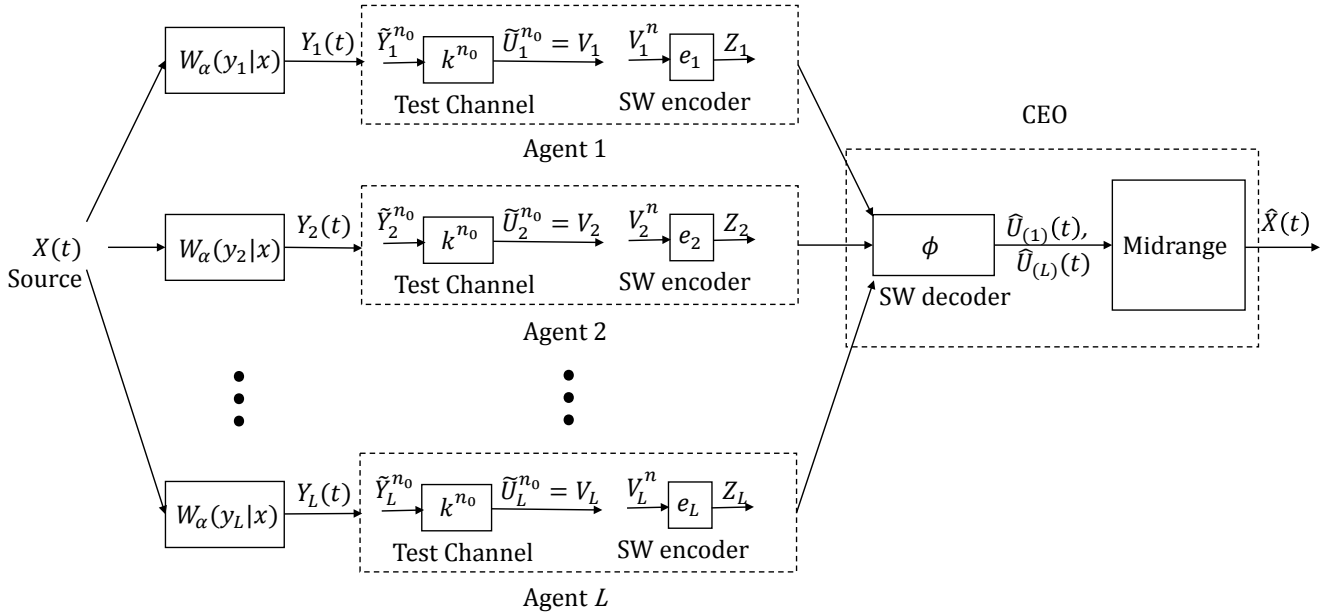


Fig. 1. A block diagram of system model for the Example 1 using the Clayton copula based observation channel $W_\alpha(y_i|x)$ given by (11) and test channel $U = Y + N$ where N is a k -peak noise. Here k^{n_0} represents the block code that approximates the test channel.

³It is straightforward to prove that the property holds even when there is overlap among the shifted versions.

III. DIRECT CODING THEOREM

The structure of the achievable scheme is a layered architecture, with scalar quantization followed by Slepian-Wolf entropy coding, just like for the Gaussian CEO problem [7] and other source coding problems [14], [29], [30]. The following are key steps of the analysis: quantization of alphabets, codes that approximate the forward test channel, Slepian-Wolf encoding and decoding, and estimation at the CEO.

Every agent uses a two-stage encoding scheme. In the first stage, a block of observations are mapped to codewords from a codebook which is identical for all agents. The second stage is an index encoder that performs Slepian-Wolf encoding of the codewords [31], [32]. For decoding, the CEO first performs index decoding to determine the L codewords corresponding to each of the agents, and then estimates the source value at each instant based on a midrange estimator [33], [34].

The key aspect of the proof is a choice of forward test channel which is characterized by an auxiliary random variable U . We choose the test channel from Y to U , denoted by $Q(u|y)$ where $U \in \mathcal{S}(W)$, so as to induce a distribution $f_{U|X}(u|x)$ which satisfies Property 1.

A. Quantization of alphabets

To design the coding scheme, we start by quantizing the continuous alphabets. Denote by \tilde{X} , \tilde{Y} , and \tilde{U} the quantized versions of random variables X , Y , and U , respectively. Their corresponding alphabets are denoted by $\tilde{\mathcal{X}}$, $\tilde{\mathcal{Y}}$, and $\tilde{\mathcal{U}}$ respectively. Conditions to be satisfied by the quantization are as follows:

$$E(U - \tilde{U})^2 \leq \delta_0 \quad (15)$$

$$|I(Y; U) - I(\tilde{Y}, \tilde{U})| \leq \delta_1 \quad (16)$$

$$|I(X; U) - I(\tilde{X}, \tilde{U})| \leq \delta_2, \quad (17)$$

where $\delta_j > 0$, for $j = 0, 1, 2$. There exists quantization schemes that achieve each of these above constraints individually: (15) from that fact that $EU^2 < \infty$, and (16) and (17) from the definition of mutual information for arbitrary ensembles [35]. Therefore, a common refinement of the quantization schemes that achieve (15)–(17) separately will satisfy them simultaneously. This quantization induces a corresponding joint probability distribution for the quantized versions \tilde{X} , \tilde{Y} , and \tilde{U} :

$$\begin{aligned} P_{\tilde{Y}, \tilde{U}}(\tilde{y}, \tilde{u}) &= \int_{\{(y,u): \text{quant}(y)=\tilde{y}, \text{quant}(u)=\tilde{u}\}} f_{Y,U}(y, u) dy du \\ P_{\tilde{X}, \tilde{U}}(\tilde{x}, \tilde{u}) &= \int_{\{(x,u): \text{quant}(x)=\tilde{x}, \text{quant}(u)=\tilde{u}\}} f_X(x) f_{U|X}(u|x) dx du \\ \tilde{W}_\alpha(\tilde{y}|x) &= \int_{\{y: \text{quant}(y)=\tilde{y}\}} W_\alpha(y|x) dy \\ Q(\tilde{u}|\tilde{y}) &= \frac{P_{\tilde{Y}, \tilde{U}}(\tilde{y}, \tilde{u})}{P_{\tilde{Y}}(\tilde{y})}. \end{aligned}$$

Note that any letters \tilde{y} of zero measure are removed from $\tilde{\mathcal{Y}}$.

B. Codes that approximate the test channel

The encoding scheme works on the quantized version \tilde{Y}_i . The basic idea is to build a block code between the quantized versions \tilde{Y}_i and \tilde{U}_i , and show that the designed block code approximates the test channel $Q(u|y)$ that arises from satisfying Property 1. Let k^{n_0} be a block code of length n_0 from $\tilde{\mathcal{Y}}^{n_0}$ to $\tilde{\mathcal{U}}^{n_0}$. This map, k^{n_0} , induces the following joint distribution between the blocks $\tilde{Y}^{n_0} = [\tilde{Y}(1), \dots, \tilde{Y}(n_0)]$ and $\tilde{U}^{n_0} = [\tilde{U}(1), \dots, \tilde{U}(n_0)]$:

$$\hat{P}^{n_0}(\tilde{Y}^{n_0} = \tilde{y}^{n_0}, \tilde{U}^{n_0} = \tilde{u}^{n_0}) = P_{\tilde{Y}^{n_0}}(\tilde{y}^{n_0}) \mathbb{1}_{\{k^{n_0}(\tilde{y}^{n_0}) = \tilde{u}^{n_0}\}},$$

where $\mathbb{1}_{\mathcal{A}}$ is the indicator function which is 1 when the event \mathcal{A} is true and 0 otherwise. Also, the corresponding marginals and conditionals are given by

$$\hat{P}(\tilde{Y}(t) = \tilde{y}, \tilde{U}(t) = \tilde{u}) = E_{P_{\tilde{Y}^{n_0}}} \mathbb{1}_{\{\tilde{U}(t) = \tilde{u}, \tilde{Y}(t) = \tilde{y}\}}$$

$$\hat{Q}(\tilde{U}(t) = \tilde{u} | \tilde{Y}(t) = \tilde{y}) = \frac{\hat{P}(\tilde{Y}(t) = \tilde{y}, \tilde{U}(t) = \tilde{u})}{P_{\tilde{Y}}(\tilde{Y}(t) = \tilde{y})}.$$

Now the existence of a block code $k^{n_0} : \tilde{\mathcal{Y}}^{n_0} \rightarrow \tilde{\mathcal{U}}^{n_0}$ which approximates a test channel $Q(u|y)$ arising from Property 1 follows from [7, Proposition 3.1], which is stated here without proof.

Proposition 2 ([7]). *For every $\epsilon_0, \delta_3 > 0$, there exists a deterministic map $k^{n_0} : \tilde{\mathcal{Y}}^{n_0} \rightarrow \tilde{\mathcal{U}}^{n_0}$ with the range cardinality M such that*

$$\frac{1}{n_0} \log M \leq I(Y; U) + \delta_3 \quad (18)$$

and

$$\sum_{\tilde{u} \in \tilde{\mathcal{U}}} |\hat{Q}(\tilde{U}(t) = \tilde{u} | x) - Q(\tilde{U}(t) = \tilde{u} | x)| \leq \frac{\epsilon_0}{|\tilde{\mathcal{X}}|}$$

for all $t = 1, \dots, n_0$ and all real x .

C. Encoding and decoding

The encoding is performed in two stages: in the first stage, the agents use the identical deterministic mapping k^{n_0} of Proposition 2 to encode their quantized observation block $\tilde{Y}_i^{n_0}$ into codewords $\tilde{U}_i^{n_0}$; and in the second stage, Slepian-Wolf encoding [31] is used to encode the index of each agent's codeword $\tilde{U}_i^{n_0}$. Let the index of codeword $\tilde{U}_i^{n_0}$ in the codebook be denoted by V_i , for $i = 1, \dots, L$. We will use the index to represent the codeword due to the one-to-one correspondence between the index and the codeword, therefore, we have $V_i = \tilde{U}_i^{n_0}$. Note that V_1, \dots, V_L are correlated and Slepian-Wolf encoding of the indices is used to remove that correlation across agents. This is done by index encoding n -length block of indices of agent i , represented as $V_i^n = [V_i(1), \dots, V_i(n)]$, where $V_i(t)$ is the t th component of the n -block of the indices of agent i . This block of indices is then mapped to a smaller index set using a mapping $e_i : \tilde{\mathcal{U}}^{n_0} \rightarrow \{0, \dots, N_i - 1\}$, for $i = 1, \dots, L$, where N_i and n are chosen to be sufficiently large to ensure a negligible decoding error. The sum rate per source symbol is given by

$$R = \frac{1}{nn_0} \sum_{i=1}^L \log N_i.$$

Therefore, we have a complete encoder $h_i = e_i \circ k^{n_0} : \tilde{\mathcal{Y}}^{n_0} \rightarrow \{0, 1, 2, \dots, N_i - 1\}$, where ' \circ ' is the composition operator. Let the output of this encoder be represented by $Z_i = h_i(\tilde{Y}_i^{n_0}) \in \{0, 1, \dots, N_i - 1\}$.

The CEO receives the indices Z_1, \dots, Z_L corresponding to the L agents. It first recovers the block of indices \hat{V}_i^n , for all i using a mapping $\phi : \prod_{i=1}^L \{0, 1, \dots, N_i - 1\} \rightarrow \prod_{i=1}^L \tilde{\mathcal{U}}_i^{n_0}$. The output of this decoder, represented as $\hat{V}_i^n = [\hat{V}_i(1), \dots, \hat{V}_i(n)] = [\hat{U}_i^{n_0}(1), \dots, \hat{U}_i^{n_0}(n)]$, is the decoded super codeword and $\hat{U}_i^{n_0}(t)$ is the decoded version of $\tilde{U}_i^{n_0}(t)$. From the Slepian-Wolf theorem (cf. [7, Proposition 3.2]), we know there exist encoders $\{e_i\}$ and a decoder ϕ such that the codewords can be recovered with negligible error probability for sufficiently large block size n .

Proposition 3 ([7]). *For every $\epsilon_1, \lambda > 0$, there exists sufficiently large L, n , and index encoders e_1, \dots, e_L and index decoder ϕ such that*

$$\frac{R}{L} \leq \frac{1}{n_0} H(\tilde{U}^{n_0} | \tilde{X}^{n_0}) + \epsilon_1, \quad (19)$$

$$\Pr\{(\hat{U}_1^{n_0}, \dots, \hat{U}_L^{n_0}) \neq (\tilde{U}_1^{n_0}, \dots, \tilde{U}_L^{n_0}) \leq \lambda\}, \quad (20)$$

where $\tilde{X}^{n_0} = [\tilde{X}(1), \dots, \tilde{X}(n_0)]$.

D. Further analysis of code rate

Note that the bound on sum rate per agent R/L in (19) is in terms of the distributions of \tilde{U} and \tilde{X} . By further analyzing the code rate, we can determine a bound which is a function of the distributions of the unquantized versions, X and U . For this we use the closeness of the marginal distribution induced by the encoding function k^{n_0} to the test channel statistics, to bound the entropy terms. Let $H(\tilde{X})$ denote the entropy of the quantized random variable \tilde{X} , then we have

$$\frac{1}{n_0}H(\tilde{U}^{n_0}|\tilde{X}^{n_0}) = \frac{1}{n_0}(H(\tilde{U}^{n_0}, \tilde{X}^{n_0}) - H(\tilde{X}^{n_0})) \quad (21)$$

$$= \frac{1}{n_0}H(\tilde{U}^{n_0}) + \frac{1}{n_0}H(\tilde{X}^{n_0}|\tilde{U}^{n_0}) - \frac{1}{n_0}H(\tilde{X}^{n_0}) \quad (22)$$

$$= \frac{1}{n_0}H(\tilde{U}^{n_0}) + \frac{1}{n_0}H(\tilde{X}^{n_0}|\tilde{U}^{n_0}) - \frac{1}{n_0}\sum_{t=1}^{n_0}H(\tilde{X}(t)) \quad (23)$$

$$\leq \frac{1}{n_0}\log M + \frac{1}{n_0}H(\tilde{X}^{n_0}|\tilde{U}^{n_0}) - \frac{1}{n_0}\sum_{t=1}^{n_0}H(\tilde{X}(t)) \quad (24)$$

$$\leq I(Y; U) + \frac{1}{n_0}H(\tilde{X}^{n_0}|\tilde{U}^{n_0}) - \frac{1}{n_0}\sum_{t=1}^{n_0}H(\tilde{X}(t)) + \delta_3 \quad (25)$$

where (23) is due to the independent nature of source $X(t)$ over time, (24) is due to the upper bound of $H(\tilde{U}^{n_0})$ by the logarithm of number of codewords M , and (25) is by (18). Next, $H(\tilde{X}^{n_0}|\tilde{U}^{n_0})$ can be further bounded as:

$$\begin{aligned} H(\tilde{X}^{n_0}|\tilde{U}^{n_0}) &= \sum_{t=1}^{n_0} H(\tilde{X}(t)|\tilde{U}^{n_0}, \tilde{X}(1), \dots, \tilde{X}(t-1)) \\ &\leq \sum_{t=1}^{n_0} H(\tilde{X}(t)|\tilde{U}(t)) \end{aligned} \quad (26)$$

using the fact conditioning only reduces entropy. Therefore, we have

$$\frac{1}{n_0}H(\tilde{U}^{n_0}|\tilde{X}^{n_0}) \leq I(Y; U) + \frac{1}{n_0}H(\tilde{X}^{n_0}|\tilde{U}^{n_0}) - \frac{1}{n_0}\sum_{t=1}^{n_0}H(\tilde{X}(t)) + \delta_3 \quad (27)$$

$$\leq I(Y; U) + \frac{1}{n_0}\sum_{t=1}^{n_0}H(\tilde{X}(t)|\tilde{U}(t)) - \frac{1}{n_0}\sum_{t=1}^{n_0}H(\tilde{X}(t)) + \delta_3$$

$$= I(Y; U) - \frac{1}{n_0}\sum_{t=1}^{n_0}I(\tilde{X}(t), \tilde{U}(t)) + \delta_3 \quad (28)$$

$$\leq I(Y; U) - I(\tilde{X}, \tilde{U}) + \delta_3 + 2\epsilon_0 \log \frac{|\tilde{\mathcal{U}}||\tilde{\mathcal{X}}|}{\epsilon_0} \quad (29)$$

where (29) is due to the t -symmetry of the encoder and [7, Proposition A.3].

Thus, using (16), we have:

$$\frac{1}{n_0}H(\tilde{U}^{n_0}|\tilde{X}^{n_0}) \leq I(Y; U) - I(\tilde{X}, \tilde{U}) + \delta_3 + 2\epsilon_0 \log \frac{|\tilde{\mathcal{U}}||\tilde{\mathcal{X}}|}{\epsilon_0} \quad (30)$$

$$\leq I(Y; U) - I(X; U) + \delta_2 + \delta_3 + 2\epsilon_0 \log \frac{|\tilde{\mathcal{U}}||\tilde{\mathcal{X}}|}{\epsilon_0} \quad (31)$$

due to (16). Due to the Markov chain relationship $X \rightarrow Y \rightarrow U$, the right hand side can be further simplified to

$$\frac{1}{n_0}H(\tilde{U}^{n_0}|\tilde{X}^{n_0}) \leq I(Y; U|X) + \delta_2 + \delta_3 + 2\epsilon_0 \log \frac{|\tilde{\mathcal{U}}||\tilde{\mathcal{X}}|}{\epsilon_0} \quad (32)$$

By choosing $\delta_2, \delta_3, \epsilon_0, \epsilon_2$ such that

$$\delta_2 + \delta_3 + 2\epsilon_0 \log \frac{|\tilde{\mathcal{U}}||\tilde{\mathcal{X}}|}{\epsilon_0} < \epsilon_2,$$

we have

$$\frac{R}{L} \leq I(Y; U|X) + \epsilon_1 + \epsilon_2. \quad (33)$$

Having determined a bound on sum rate, the next step is to bound the minimum quadratic distortion.

E. Estimation scheme

The CEO, after decoding the codewords sent by the agents $(\hat{U}_1^{n_0}, \dots, \hat{U}_L^{n_0})$, estimates the source $X(t)$ on an instant-by-instant basis. Since the range of $U_i(t)$ depends on $X(t)$, we first estimate the midrange of data $\hat{U}_1(t), \dots, \hat{U}_L(t)$. The midrange estimator [33], [34], [36], [37] is the maximally efficient estimator for the center of a uniform distribution. The midrange estimator also seems to work well for estimating the location parameter of other distributions of bounded support and it is more effective than the sample mean for many distributions such as the cosine distribution, parabolic distribution, rectangular distribution, and inverted parabolic distribution [37], though the best estimator depends on the distribution of the source that is to be estimated. For these reasons, the midrange estimator is used in this paper.

After estimating the midrange of data, using the inverse function $l(\cdot)$ as follows (cf. Property 1),

$$\hat{X}(t) = 2l\left(\frac{\hat{U}_{(1)}(t) + \hat{U}_{(L)}(t)}{2}\right), \quad (34)$$

we get an estimate of $X(t)$. Here $\hat{U}_{(i)}(t)$ are the order statistics of $\hat{U}_i(t)$ [38]. Note that $E\left[\frac{\hat{U}_{(1)} + \hat{U}_{(L)}}{2}\right] = E\left[\frac{a(X) + b(X)}{2}\right]$.

We can now derive an upper bound on the distortion, following a method similar to Akay, et al. [39]:

$$\begin{aligned} E[\hat{X}(t) - X(t)]^2 &\leq K^2 E\left[\left(\frac{\hat{U}_{(1)}(t) + \hat{U}_{(L)}(t)}{2} - \frac{a(X(t)) + b(X(t))}{2}\right)^2\right] \\ &= K^2 E_X E_{U|X}\left[\left(\frac{\hat{U}_{(1)}(t) + \hat{U}_{(L)}(t)}{2} - \frac{a(X(t)) + b(X(t))}{2}\right)^2 \middle| X(t)\right] \end{aligned} \quad (35)$$

$$\leq 2K^2 E_X E_{U|X}\left[\left(\frac{U_{(1)}(t) + U_{(L)}(t)}{2} - \frac{a(X(t)) + b(X(t))}{2}\right)^2 \middle| X(t)\right] + \epsilon_3 \quad (36)$$

where $U_{(i)}(t)$ are the order statistics of $U_i(t)$; the first inequality is due to Lipschitz continuity of the function $l(\cdot)$ with Lipschitz constant K , and (36) follows from Proposition 4 in the Appendix.

Now we evaluate the main term in (36); for notational simplicity, we drop the dependence on t and the dependence of $a(\cdot)$ and $b(\cdot)$ on X . However, we need to be aware of the dependence of the limits a and b on the unknown X . As $f_{U|X}(u|x)$ does not vanish at the endpoints, there exists ϵ and δ such that $f_{U|X}(u|x) \geq \delta$ for $a \leq u \leq a + \epsilon$ and $b - \epsilon \leq u \leq b$. Now,

$$\begin{aligned} E_{U|X}\left[\left(\frac{U_{(1)} + U_{(L)}}{2} - \frac{a + b}{2}\right)^2 \middle| X\right] &= \int_{u_{(1)} > a + \epsilon \text{ or } u_{(L)} < b - \epsilon} \left(\frac{u_{(1)} + u_{(L)}}{2} - \frac{a + b}{2}\right)^2 f_{u_{(1)}, u_{(L)}|X} du_{(1)} du_{(L)} \\ &\quad + \int_{u_{(1)} < a + \epsilon \text{ and } u_{(L)} > b - \epsilon} \left(\frac{u_{(1)} + u_{(L)}}{2} - \frac{a + b}{2}\right)^2 f_{u_{(1)}, u_{(L)}|X} du_{(1)} du_{(L)}. \end{aligned} \quad (37)$$

Since $(u_{(1)} + u_{(L)})/2 \in [a, b]$, we have:

$$\left(\frac{u_{(1)} + u_{(L)}}{2} - \frac{a + b}{2}\right)^2 \leq \left(\frac{b - a}{2}\right)^2$$

and the first term on right side of (37) can be bounded as:

$$\begin{aligned} & \int_{u_{(1)} > a + \epsilon \text{ or } u_{(L)} < b - \epsilon} \left(\frac{u_{(1)} + u_{(L)}}{2} - \frac{a + b}{2} \right)^2 f_{u_{(1)}, u_{(L)} | X} du_{(1)} du_{(L)} \\ & \leq 2^{-2} (b - a)^2 \Pr \{ u_{(1)} > a + \epsilon \text{ or } u_{(L)} < b - \epsilon | X \}. \end{aligned}$$

Since the $\{U_i\}$ are conditionally independent given X , we can simplify this further as:

$$\Pr \{ u_{(1)} > a + \epsilon \text{ or } u_{(L)} < b - \epsilon | X \} \leq \Pr \{ u_{(1)} > a + \epsilon | X \} + \Pr \{ u_{(L)} < b - \epsilon | X \} \quad (38)$$

$$= \prod_{i=1}^L \Pr \{ u_i > a + \epsilon | X \} + \prod_{i=1}^L \Pr \{ u_i < b - \epsilon | X \} \quad (39)$$

$$= \prod_{i=1}^L (1 - \Pr \{ u_i \leq a + \epsilon | X \}) + \prod_{i=1}^L (1 - \Pr \{ u_i \geq b - \epsilon | X \}). \quad (40)$$

Since, $f_{U|X}(u|x) \geq \delta$ for $a \leq u \leq a + \epsilon$ and $b - \epsilon \leq u \leq b$, $\Pr \{ u_i \leq a + \epsilon | X \} \geq \delta \epsilon$ and $\Pr \{ u_i \geq b - \epsilon | X \} \geq \delta \epsilon$. Therefore,

$$\begin{aligned} \Pr \{ u_{(1)} > a + \epsilon \text{ or } u_{(L)} < b - \epsilon | X \} & \leq \prod_{i=1}^L (1 - \Pr \{ u_i \leq a + \epsilon | X \}) + \prod_{i=1}^L (1 - \Pr \{ u_i \geq b - \epsilon | X \}) \\ & \leq \prod_{i=1}^L (1 - \delta \epsilon) + \prod_{i=1}^L (1 - \delta \epsilon) \end{aligned} \quad (41)$$

$$\leq 2(1 - \delta \epsilon)^L. \quad (42)$$

To evaluate the second term in the right side of (37), we define the following variables:

$$\xi = L(1 - F_{U|X}(u_{(L)})), \quad b - \epsilon \leq u_{(L)} \leq b \quad (43)$$

$$\eta = LF_{U|X}(u_{(1)}), \quad a \leq u_{(1)} \leq a + \epsilon, \quad (44)$$

where $F_{U|X}$ is the conditional cumulative distribution function of U given X . These variables have the following marginal and joint densities [39]:

$$f_{\xi}(s) = f_{\eta}(s) = \left(1 - \frac{s}{L}\right)^{L-1}, \quad 0 \leq s \leq L \quad (45)$$

$$f_{\xi, \eta}(s_1, s_2) = \frac{L-1}{L} \left(1 - \frac{s_1 + s_2}{L}\right)^{L-2}, \quad s_1, s_2 \geq 0, \quad s_1 + s_2 \leq L.$$

Also, as $L \rightarrow \infty$, ξ and η become independent and $f_{\xi}(s), f_{\eta}(s) \rightarrow e^{-s}$.

From the above definitions, we have

$$\xi = L \int_{u_{(L)}}^b f_{u|x} du \geq \delta L(b - u_{(L)}), \quad (46)$$

$$\eta = L \int_a^{u_{(1)}} f_{u|x} du \geq \delta L(u_{(1)} - a), \quad (47)$$

provided $u_{(1)} \leq a + \epsilon$ and $b - \epsilon \leq u_{(L)}$. Therefore, for the second term, we have

$$\left| \frac{u_{(1)} + u_{(L)}}{2} - \frac{a + b}{2} \right|^2 = \frac{1}{4} |(u_{(1)} - a) - (b - u_{(L)})|^2 \quad (48)$$

$$\leq \frac{1}{4} [|u_{(1)} - a|^2 + |b - u_{(L)}|^2] \quad (49)$$

$$\leq \frac{\xi^2 + \eta^2}{4\delta^2 L^2}, \quad (50)$$

where we used the fact that $|A - B|^2 \leq A^2 + B^2$ for $A, B > 0$.

Now using the inequalities we have developed, we can bound the distortion in (36) as:

$$D(L, R) \leq 2K^2 E_X \left[\frac{(b-a)^2(1-\delta\epsilon)^L}{2} + \frac{1}{4\delta^2 L^2} \int_0^{L(1-F_{U|X}(b-\epsilon))} \int_0^{LF_{U|X}(a+\epsilon)} (s_1^2 + s_2^2) f_{\xi,\eta}(s_1, s_2) ds_1 ds_2 \right] + \epsilon_3. \quad (51)$$

Using (33) and (51), we get:

$$R^2 D(L, R) \leq L^2 I^2(Y; U|X) \left(2K^2 E_X \left[\frac{(b-a)^2(1-\delta\epsilon)^L}{2} + \frac{1}{4\delta^2 L^2} \int_0^{L(1-F_{U|X}(b-\epsilon))} \int_0^{LF_{U|X}(a+\epsilon)} (s_1^2 + s_2^2) f_{\xi,\eta}(s_1, s_2) ds_1 ds_2 \right] + \epsilon_3 \right).$$

By taking limits $L, R \rightarrow \infty$, we have:

$$\begin{aligned} \beta(\alpha) &= \lim_{L, R \rightarrow \infty} R^2 D(L, R) \\ &\leq I^2(Y; U|X) \left(2K^2 E_X \left[\frac{1}{4\delta^2} \int_0^\infty \int_0^\infty (s_1^2 + s_2^2) \lim_{L \rightarrow \infty} f_{\xi,\eta}(s_1, s_2) ds_1 ds_2 \right] \right) \end{aligned} \quad (52)$$

$$\begin{aligned} &= I^2(Y; U|X) \left(2K^2 E_X \left[\frac{1}{4\delta^2} \int_0^\infty \int_0^\infty (s_1^2 + s_2^2) e^{-s_1} e^{-s_2} ds_1 ds_2 \right] \right) \\ &= I^2(Y; U|X) \left(2K^2 E_X \left[\frac{1}{2\delta^2} \int_0^\infty s^2 e^{-s} ds \right] \right) \end{aligned} \quad (53)$$

$$= \frac{2K^2}{\delta^2} I^2(Y; U|X) > 0 \quad (54)$$

where U is chosen to satisfy Property 1 and $K > 0$ is a constant. Therefore, we have

$$\beta(\alpha) \leq \frac{2K^2}{\delta^2} \left(\min_{U \in \mathcal{S}(W)} I(Y; U|X) \right)^2. \quad (55)$$

This concludes the achievability proof. Note that the bound only depends on the conditional mutual information $I(Y; U|X)$ which corresponds to the compression of the observation noise. The compression of the source X does not appear in the bound, since such a term vanishes because the number of agents L grows without bound.

IV. CONVERSE CODING THEOREM

The converse for the quadratic non-regular CEO problem is similar in structure to the converse for the quadratic Gaussian CEO [7] and the discrete CEO problem [6]. The proof uses a lower bound on the distortion function similar to the Bayesian Cramér-Rao lower bound used in [7]. However, note that the source distribution herein does not satisfy the regularity conditions required for using the Cramér-Rao bound [25]. Therefore, we use a version of the extended Chazan-Zakai-Ziv bound [26], [40], [41] which is first stated here without proof.

Lemma 1. *For estimating a random scalar parameter $x \sim f_X(x)$ with support on $[0, T]$ using data $\mathbf{z} = [z_1, \dots, z_k]$ with conditional distribution $f(\mathbf{z}|x)$, the MSE between x and $\hat{x}(\mathbf{z})$ is bounded as follows:*

$$E(x - \hat{x}(\mathbf{z}))^2 \geq \frac{1}{2T} \int_{h=0}^T h \left[\int_{\theta=0}^{T-h} (f_X(\theta) + f_X(\theta+h)) P_{\min}(\theta, \theta+h) d\theta \right] dh, \quad (56)$$

where $P_{\min}(\theta, \theta+h)$ is the minimum error probability corresponding to the following binary hypothesis testing problem:

$$\begin{aligned} H_0 &: \mathbf{z} \sim f(\mathbf{z}|x), \quad x = \theta, & \Pr(H_0) &= \frac{f_X(\theta)}{f_X(\theta) + f_X(\theta+h)}, \\ H_1 &: \mathbf{z} \sim f(\mathbf{z}|x), \quad x = \theta+h, & \Pr(H_1) &= \frac{f_X(\theta+h)}{f_X(\theta) + f_X(\theta+h)}. \end{aligned}$$

The above Chazan-Zakai-Ziv bound falls under the family of Ziv-Zakai bounds. Ziv-Zakai bounds have been shown to be useful bounds for all regions of operation unlike other bounds (for example, Cramér-Rao bound) that have limited applicability [41]. This family of bounds build on the original Ziv-Zakai bound [42] and have the advantage of being independent of bias and very tight in most cases. A detailed study of this family of bounds can be found in [41].

Note that this lemma bounds the performance of an estimation problem in terms of the performance of a sequence of detection problems. Therefore, as we shall see, we get Chernoff information rather than Fisher information as seen in the estimation problem in the quadratic Gaussian CEO [7].

Using Lemma 1, we now prove our converse. Let $\{C_i^n\}_{i=1}^L$ be L codes of block length n , corresponding to the L agents, with respective rates R_1, R_2, \dots, R_L . We use the genie-aided approach to determine the lower bound as follows: Let the CEO implement n estimators O_t for $t = 1, \dots, n$ where O_t estimates $X(t)$ given all components of the source word x^n except $x(t)$. Recall the definition of $X^n = [X(1), \dots, X(n)]$ and further define $Y_i^n = [Y_i(1), \dots, Y_i(n)]$. We have

$$\begin{aligned} nR_i &= \log |\mathcal{C}_i^n| \\ &\geq I(Y_i^n; C_i | X^n) \\ &= \sum_{t=1}^n I(Y_i(t); C_i | Y_i^{t-1}, X^n) \end{aligned} \quad (57)$$

$$\begin{aligned} &= \sum_{t=1}^n [h(Y_i(t) | Y_i^{t-1}, X^n) - h(Y_i(t) | C_i, Y_i^{t-1}, X^n)] \\ &= \sum_{t=1}^n [h(Y_i(t) | X^n) - h(Y_i(t) | C_i, Y_i^{t-1}, X^n)] \end{aligned} \quad (58)$$

$$\begin{aligned} &\geq \sum_{t=1}^n [h(Y_i(t) | X^n) - h(Y_i(t) | C_i, X^n)] \\ &= \sum_{t=1}^n I(Y_i(t); C_i | X^n), \end{aligned} \quad (59)$$

where X is the generic source random variable, Y_i is the noisy version of X as observed by agent i ; (57) is from the product rule of mutual information, (58) is due to the independence of $Y(t)$ across time, and (59) follows since conditioning only reduces entropy.

Hence, we get a lower bound on the sum rate R as follows:

$$R \geq \frac{1}{n} \sum_{t=1}^n \sum_{i=1}^L I(Y_i(t); C_i | X^n).$$

Define $\check{X}_t = (X_1, \dots, X_{t-1}, X_{t+1}, \dots, X_n)$ and let $U_i(t, \check{x}_t)$ be a random variable whose joint distribution with $X(t)$ and $Y_i(t)$ is:

$$\begin{aligned} &\Pr \{x \leq X(t) \leq x + dx, y \leq Y_i(t) \leq y + dy, U_i(t, \check{x}_t) = c\} \\ &= f_X(x) W_\alpha(y|x) \Pr(C_i = c | Y_i(t) = y, X(t) = x, \check{X}_t = \check{x}_t) dx dy \\ &= f_X(x) W_\alpha(y|x) \Pr(C_i = c | Y_i(t) = y, \check{X}_t = \check{x}_t) dx dy, \end{aligned}$$

since the codeword C_i depends on $X(t)$ only through $Y_i(t)$. Therefore, for each i and any fixed \check{x}_t , we have the Markov chain relationship $X(t) \rightarrow Y_i(t) \rightarrow U_i(t, \check{x}_t)$. Now, we can express the lower bound on R as

$$R \geq \frac{1}{n} \sum_{t=1}^n \sum_{i=1}^L E_{\check{X}_t} I(Y_i(t); U_i(t, \check{X}_t) | X(t)). \quad (60)$$

Note that in order to find a lower bound on $\beta(\alpha)$, we consider the best case where the CEO knows C_1, \dots, C_L and \check{x}_t , i.e., the CEO uses an estimator $\hat{X}(C_1, \dots, C_L, \check{x}_t)$. Using the Chazan-Zakai-Ziv bound (Lemma 1), we

have:

$$E(X(t) - \hat{X}_t)^2 \geq \frac{1}{2} \int_{h=0}^1 h \left[\int_{\theta=0}^{1-h} (f_X(\theta) + f_X(\theta + h)) P_{min,t}(\theta, \theta + h) d\theta \right] dh \quad (61)$$

where $P_{min,t}(\theta, \theta + h)$ is the minimum achievable error probability, using data $Y_1(t), \dots, Y_L(t)$ from the L agents, to differentiate between $X(t) = \theta$ and $X(t) = \theta + h$.

Therefore, from the definition of $D(L, R)$, we have:

$$\begin{aligned} D(L, R) &= \frac{1}{n} \sum_{t=1}^n E(X(t) - \hat{X}_t)^2 \\ &\geq \frac{1}{2n} \sum_{t=1}^n \left[\int_{h=0}^1 h \left[\int_{\theta=0}^{1-h} (f_X(\theta) + f_X(\theta + h)) P_{min,t}(\theta, \theta + h) d\theta \right] dh \right] \\ &\geq \frac{1}{2nL^2} \sum_{t=1}^n \left[\int_{h=0}^1 hL \left[\int_{\theta=0}^{1-h} (f_X(\theta) + f_X(\theta + h)) P_{min,t}(\theta, \theta + h) d\theta \right] d(hL) \right], \end{aligned}$$

where we have multiplied and divided the right side by L^2 . Now, using a change of variables $\tilde{h} = hL$, we get:

$$\begin{aligned} D(L, R) &\geq \frac{1}{2nL^2} \sum_{t=1}^n \int_{(hL)=0}^L (hL) \int_{\theta=0}^{1-(hL)/L} (f_X(\theta) + f_X(\theta + (hL)/L)) P_{min,t}(\theta, \theta + (hL)/L) d\theta d(hL) \\ &= \frac{1}{2nL^2} \sum_{t=1}^n \left[\int_{\tilde{h}=0}^L \tilde{h} \left[\int_{\theta=0}^{1-\tilde{h}/L} (f_X(\theta) + f_X(\theta + \tilde{h}/L)) P_{min,t}(\theta, \theta + \tilde{h}/L) d\theta \right] d\tilde{h} \right] \\ &\geq \frac{1}{2L^2} \frac{1}{\frac{1}{n} \sum_{t=1}^n \frac{1}{\left[\int_{\tilde{h}=0}^L \tilde{h} \left[\int_{\theta=0}^{1-\tilde{h}/L} (f_X(\theta) + f_X(\theta + \tilde{h}/L)) P_{min,t}(\theta, \theta + \tilde{h}/L) d\theta \right] d\tilde{h} \right]}}, \quad (62) \end{aligned}$$

where the last step is due to the inequality of arithmetic and harmonic means.

Note that, although not explicit, $D(L, R)$ does depend on R . This dependence is implicitly visible via n (see (60)). Therefore, as can be observed below in (63), the product of R^2 and $D(L, R)$ results in a positive constant that is independent of R and does not vanish as $L \rightarrow \infty$.

Using (60) and (62), we have the following expression:

$$\begin{aligned} R^2 D(L, R) &\geq \frac{1}{n^2} \frac{n}{2L^2} \frac{\left(\sum_{t=1}^n \sum_{i=1}^L E_{\check{X}_t} I(Y_i(t); U_i(t, \check{X}_t) | X(t)) \right)^2}{\sum_{t=1}^n \left[\int_{\tilde{h}=0}^L \tilde{h} \left[\int_{\theta=0}^{1-\tilde{h}/L} (f_X(\theta) + f_X(\theta + \tilde{h}/L)) P_{min,t}(\theta, \theta + \tilde{h}/L) d\theta \right] d\tilde{h} \right]^{-1}} \\ &= \frac{1}{2nL^2} \frac{\left(\sum_{t=1}^n \sum_{i=1}^L E_{\check{X}_t} I(Y_i(t); U_i(t, \check{X}_t) | X(t)) \right)^2}{\sum_{t=1}^n \left[\int_{\tilde{h}=0}^L \tilde{h} \left[\int_{\theta=0}^{1-\tilde{h}/L} (f_X(\theta) + f_X(\theta + \tilde{h}/L)) P_{min,t}(\theta, \theta + \tilde{h}/L) d\theta \right] d\tilde{h} \right]^{-1}} \\ &= \frac{1}{2nL^2} \frac{\sum_{t=1}^n \sum_{t'=1}^n \sum_{i=1}^L \sum_{i'=1}^L E_{\check{X}_t} I(Y_i(t); U_i(t, \check{X}_t) | X(t)) E_{\check{X}_{t'}} I(Y_{i'}(t'); U_{i'}(t', \check{X}_{t'}) | X(t'))}{\sum_{t=1}^n \left[\int_{\tilde{h}=0}^L \tilde{h} \left[\int_{\theta=0}^{1-\tilde{h}/L} (f_X(\theta) + f_X(\theta + \tilde{h}/L)) P_{min,t}(\theta, \theta + \tilde{h}/L) d\theta \right] d\tilde{h} \right]^{-1}} \\ &\geq \min_t \frac{1}{2nL^2} \frac{\sum_{t'=1}^n \sum_{i=1}^L \sum_{i'=1}^L E_{\check{X}_t} I(Y_i(t); U_i(t, \check{X}_t) | X(t)) E_{\check{X}_{t'}} I(Y_{i'}(t'); U_{i'}(t', \check{X}_{t'}) | X(t'))}{\left[\int_{\tilde{h}=0}^L \tilde{h} \left[\int_{\theta=0}^{1-\tilde{h}/L} (f_X(\theta) + f_X(\theta + \tilde{h}/L)) P_{min,t}(\theta, \theta + \tilde{h}/L) d\theta \right] d\tilde{h} \right]^{-1}} \\ &\geq \min_{t, t', i, i'} E_{\check{X}_t} I(Y_i(t); U_i(t, \check{X}_t) | X(t)) E_{\check{X}_{t'}} I(Y_{i'}(t'); U_{i'}(t', \check{X}_{t'}) | X(t')) \\ &\quad \times \int_{\tilde{h}=0}^L \frac{\tilde{h}}{2} \left[\int_{\theta=0}^{1-\tilde{h}/L} (f_X(\theta) + f_X(\theta + \tilde{h}/L)) P_{min,t}(\theta, \theta + \tilde{h}/L) d\theta \right] d\tilde{h} \quad (63) \end{aligned}$$

where Proposition 5 from the Appendix is used for the last two inequalities. Since the input sequence $X(t)$ is i.i.d. over time, the minimum for the ‘primed’ variables and the ‘unprimed’ variables is the same. Therefore, we can

further simplify the inequality in (63) as:

$$R^2 D(L, R) \geq \left(\min_{t,i} E_{\check{X}_t} I(Y_i(t); U_i(t, \check{X}_t) | X(t)) \right)^2 \int_{\tilde{h}=0}^L \frac{\tilde{h}}{2} \left[\int_{\theta=0}^{1-\tilde{h}/L} (f_X(\theta) + f_X(\theta + \tilde{h}/L)) P_{min,t}(\theta, \theta + \tilde{h}/L) d\theta \right] d\tilde{h}.$$

Further simplification gives the following

$$R^2 D(L, R) \geq \left(\min_{t,i,\check{X}_t} I(Y_i(t); U_i(t, \check{X}_t) | X(t)) \right)^2 \int_{\tilde{h}=0}^L \frac{\tilde{h}}{2} \left[\int_{\theta=0}^{1-\tilde{h}/L} (f_X(\theta) + f_X(\theta + \tilde{h}/L)) P_{min,t}(\theta, \theta + \tilde{h}/L) d\theta \right] d\tilde{h}.$$

Now as $L \rightarrow \infty$, using the Chernoff-Stein Lemma [43], the error probability $P_{min,t}(\theta, \theta + \tilde{h}/L)$ is given as $e^{-LC(\theta, \theta + \tilde{h}/L)}$ where $G_\theta(\tilde{h}/L) \triangleq \mathcal{C}(\theta, \theta + \tilde{h}/L)$ is the Chernoff information between the conditional densities of y given $x = \theta$ and $x = \theta + \tilde{h}/L$. It is given by the following

$$G_\theta(\tilde{h}/L) = -\min_s \log \left(\int W_\alpha^s(y|\theta) W_\alpha^{1-s}(y|\theta + \tilde{h}/L) dy \right).$$

Since the argument of $G_\theta(\tilde{h}/L)$ is close to zero as $L \rightarrow \infty$, using the Taylor expansion of $G_\theta(\Delta)$ around zero, we get:

$$G_\theta(\Delta) = G_\theta(0) + \Delta G'_\theta(\Delta)|_{\Delta=0} + O(\Delta^2). \quad (64)$$

Using this expansion, we have:

$$e^{-LC(\theta, \theta + \tilde{h}/L)} = e^{-LG_\theta(\tilde{h}/L)} \quad (65)$$

$$\begin{aligned} &= e^{-L(G_\theta(0) + \tilde{h}/L G'_\theta(\Delta)|_{\Delta=0} + O(L^{-2}))} \\ &= e^{-\tilde{h} G'_\theta(\Delta)|_{\Delta=0} + O(L^{-1})}, \end{aligned} \quad (66)$$

since $G_\theta(0) = \mathcal{C}(\theta, \theta) = 0$. Therefore,

$$\begin{aligned} \lim_{L \rightarrow \infty} R^2 D(L, R) &\geq \\ \lim_{L \rightarrow \infty} \left(\min_{t,i,\check{X}_t} I(Y_i(t); U_i(t, \check{X}_t) | X(t)) \right)^2 \int_{\tilde{h}=0}^L \frac{\tilde{h}}{2} \left[\int_{\theta=0}^{1-\tilde{h}/L} (f_X(\theta) + f_X(\theta + \tilde{h}/L)) e^{-\tilde{h} G'_\theta(\Delta)|_{\Delta=0} + O(L^{-1})} d\theta \right] d\tilde{h} \end{aligned}$$

which implies

$$\beta(\alpha) = \lim_{L,R \rightarrow \infty} R^2 D(L, R) \geq \left(\min_{U: X \rightarrow Y \rightarrow U} I(Y; U | X) \right)^2 \int_{h=0}^{\infty} h \int_{\theta=0}^1 f_X(\theta) e^{-hg(\theta)} d\theta dh,$$

where $g(\theta)$ is the first derivative of Chernoff information between the conditional densities (W_α) of the observation given $x = \theta$ and $x = \theta + \Delta$, evaluated at $\Delta = 0$ and is given by:

$$g(\theta) = \left\{ \frac{d}{d\Delta} - \left[\min_s \log \left(\int W_\alpha^s(y|\theta) W_\alpha^{1-s}(y|\theta + \Delta) dy \right) \right] \right\}_{\Delta=0}. \quad (67)$$

This concludes the converse proof.

V. DISCUSSION

We considered the non-regular CEO problem, which addresses the practical case where multiple subordinates send quantal grades of their noisy beliefs to the CEO. When the source distribution does not satisfy the regularity conditions, we get an intermediate regime of performance between the discrete CEO problem [6] and the quadratic Gaussian CEO problem [7]. A key observation is the rate of convergence depends on Chernoff information. The result expands the literature on CEO problem from the traditional case of Gaussian source distribution and Gaussian channel noise to non-regular source distributions. While the proofs are similar in structure to the traditional CEO problems, they use different techniques, which can also be applied to other non-Gaussian non-regular multiterminal

source coding problems. Our results indicate that one can expect a change in behavior for other multiterminal source coding problems as well, when the source follows non-Gaussian non-regular distribution.

There are a number of interesting future directions for research. In this work, we considered only scaling behavior of quadratic non-regular CEO problem. It is desired to derive precise characterizations for sum rate distortion and for full rate-distortion for this non-regular CEO problem as obtained by Oohama [8] and Prabhakaran, et al. [10], respectively for the quadratic Gaussian CEO problem. Similar to other CEO problems, we can observe the difference in decay rates for distortion between our result and the centralized case when agents can convene. When agents are allowed to convene, the setup is the single-terminal compression problem whose rate-distortion function under MSE was determined by Fix [28]. However, it has no simple expression and the optimizing solution has support on finite number of mass points. On the other hand, for absolute error distortion measure, rate-distortion function exists in closed form for uniform source [44] and it would be interesting to analyze the uniform CEO problem under the absolute error distortion. Gastpar and Eswaran [16] have addressed the CEO problem for non-Gaussian sources, but have considered the additive Gaussian noise channel. An interesting variant is when the source follows a regular distribution with a finite support and the measurement noise is modeled using copula. For example, beta distribution satisfies the regularity conditions and has a finite support. Also, for distributions such as cosine, parabolic, and inverted parabolic, midrange (similar to the one used in this paper) is more efficient than mean [37]. In such cases, it is interesting to explore if the minimum achievable square distortion would still exhibit a $1/R$ convergence behavior.

ACKNOWLEDGMENT

The authors would like to thank Prof. Vivek K Goyal for discussions on midrange estimators, and Prof. Pramod K. Varshney for his support during this project. We would also like to thank the anonymous reviewers for their valuable comments and suggestions that helped us improve the paper.

REFERENCES

- [1] R. B. Nelsen, *An Introduction to Copulas*. New York: Springer, 2006.
- [2] C. Genest and J. MacKay, "The joy of copulas: Bivariate distributions with uniform marginals," *Am. Stat.*, vol. 40, no. 4, pp. 280–283, Nov. 1986.
- [3] U. Cherubini, E. Luciano, and W. Vecchiato, *Copula Methods in Finance*. New York: Wiley, 2004.
- [4] K. Ratakonda, R. Williams, J. Bisceglia, R. W. Taylor, and J. Graham, "Identifying trouble patterns in complex IT services engagements," *IBM J. Res. Develop.*, vol. 54, no. 2, p. 5, Mar.-Apr. 2010.
- [5] J. B. Predd, R. Seiringer, E. H. Lieb, D. N. Osherson, H. V. Poor, and S. R. Kulkarni, "Probabilistic coherence and proper scoring rules," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4786–4792, Oct. 2009.
- [6] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.
- [7] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1549–1559, Sep. 1997.
- [8] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1057–1070, May 1998.
- [9] —, "Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2577–2593, Jul. 2005.
- [10] V. Prabhakaran, D. Tse, and K. Ramachandran, "Rate region of the quadratic Gaussian CEO problem," in *Proc. 2004 IEEE Int. Symp. Inf. Theory*, June-July 2004, p. 117.
- [11] Y. Yang and Z. Xiong, "On the generalized Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3350–3372, Jun. 2012.
- [12] S. Tavildar and P. Viswanath, "On the sum-rate of the vector Gaussian CEO problem," in *Conf. Rec. 39th Asilomar Conf. Signals, Syst. Comput.*, Oct. 2005, pp. 3–7.
- [13] J. Chen and J. Wang, "On the vector Gaussian CEO problem," in *Proc. 2011 IEEE Int. Symp. Inf. Theory*, Jul. 2011, pp. 2050–2054.
- [14] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic Gaussian two-encoder source-coding problem," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1938–1961, Sep. 2008.
- [15] J. Chen, X. Zhang, T. Berger, and S. B. Wicker, "An upper bound on the sum-rate distortion function and its corresponding rate allocation schemes for the CEO problem," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 977–987, Aug. 2004.
- [16] M. Gastpar and K. Eswaran, "On the quadratic AWGN CEO problem and non-Gaussian sources," in *Proc. 2005 IEEE Int. Symp. Inf. Theory*, Sep. 2005, pp. 219–223.
- [17] R. Soundararajan, A. B. Wagner, and S. Vishwanath, "Sum rate of the vacationing-CEO problem," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6304–6319, Oct. 2012.
- [18] P. Viswanath, "Sum rate of a class of Gaussian multiterminal source coding problems," in *Advances in Network Information Theory*, P. Gupta, G. Kramer, and A. J. van Wijngaarden, Eds. Providence: DIMACS, American Mathematical Society, 2004, pp. 43–64.
- [19] C. Hildreth, "Bayesian statisticians and remote clients," *Econometrica*, vol. 31, no. 3, pp. 422–438, Jul. 1963.

- [20] G. Kramer and S. A. Savari, “Communicating probability distributions,” *IEEE Trans. Inf. Theory*, vol. 53, no. 2, pp. 518–525, Feb. 2007.
- [21] K. R. Varshney and L. R. Varshney, “Quantization of prior probabilities for hypothesis testing,” *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4553–4562, Oct. 2008.
- [22] J. B. Rhim, L. R. Varshney, and V. K. Goyal, “Quantization of prior probabilities for collaborative distributed hypothesis testing,” *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4537–4550, Sep. 2012.
- [23] K. R. Varshney and L. R. Varshney, “Optimal grouping for group minimax hypothesis testing,” *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6511–6521, Oct. 2014.
- [24] T. S. Han and S.-I. Amari, “Statistical inference under multiterminal data compression,” *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2300–2324, Oct. 1998.
- [25] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. John Wiley & Sons, 1968.
- [26] D. Chazan, M. Zakai, and J. Ziv, “Improved lower bounds on signal parameter estimation,” *IEEE Trans. Inf. Theory*, vol. IT-21, no. 1, pp. 90–93, Jan. 1975.
- [27] H. L. Van Trees and K. L. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*. John Wiley & Sons, 2007.
- [28] S. L. Fix, “Rate distortion functions for squared error distortion measures,” in *Proc. 16th Annu. Allerton Conf. Commun. Control Comput.*, Oct. 1978, pp. 704–711.
- [29] R. Zamir and T. Berger, “Multiterminal source coding with high resolution,” *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 106–117, Jan. 1999.
- [30] S. D. Servetto, “Achievable rates for multiterminal source coding with scalar quantizers,” in *Conf. Rec. 39th Asilomar Conf. Signals, Syst. Comput.*, Oct. 2005, pp. 1762–1766.
- [31] D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources,” *IEEE Trans. Inf. Theory*, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.
- [32] T. M. Cover, “A proof of the data compression theorem of Slepian and Wolf for ergodic sources,” *IEEE Trans. Inf. Theory*, vol. IT-21, no. 2, pp. 226–228, Mar. 1975.
- [33] J. Neyman and E. S. Pearson, “On the use and interpretation of certain test criteria for purposes of statistical inference: Part I,” *Biometrika*, vol. 20A, no. 1/2, pp. 175–240, Jul. 1928.
- [34] G. R. Arce and S. A. Fontana, “On the midrange estimator,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 6, pp. 920–922, Jun. 1988.
- [35] R. L. Dobrushin, “General formulation of Shannon’s main theorem of information theory,” *Usp. Math. Nauk.*, vol. 14, no. 6(90), pp. 3–104, 1959, translated in *Am. Math. Soc. Trans.*, vol. 33, pp. 323–348.
- [36] H. Cramér, *Mathematical Methods of Statistics*. Princeton, NJ: Princeton Univ. Press, 1946.
- [37] P. R. Rider, “The midrange of a sample as an estimator of the population midrange,” *J. Am. Stat. Assoc.*, vol. 52, no. 280, pp. 537–542, Dec. 1957.
- [38] H. A. David and H. N. Nagaraja, *Order Statistics*, 3rd ed. Hoboken, NJ: Wiley-Interscience, 2003.
- [39] H. Akçay, H. Hjalmarsson, and L. Ljung, “On the choice of norms in system identification,” *IEEE Trans. Autom. Control*, vol. 41, no. 9, pp. 1367–1372, Sep. 1996.
- [40] K. L. Bell, Y. Steinberg, Y. Ephraim, and H. L. Van Trees, “Extended Ziv-Zakai lower bound for vector parameter estimation,” *IEEE Trans. Inf. Theory*, vol. 43, no. 2, pp. 624–637, Mar. 1997.
- [41] K. L. Bell, “Performance bounds in parameter estimation with application to bearing estimation,” Ph.D. dissertation, George Mason University, 1995.
- [42] J. Ziv and M. Zakai, “Some lower bounds on signal parameter estimation,” *IEEE Trans. Inf. Theory*, vol. IT-15, no. 3, pp. 386–391, May 1969.
- [43] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, 1991.
- [44] K. Yao and H. H. Tan, “Absolute error rate-distortion functions for sources with constrained magnitude,” *IEEE Trans. Inf. Theory*, vol. IT-24, no. 4, pp. 499–503, Jul. 1978.

APPENDIX

Proposition 4.

$$K^2 E \left[\left(\frac{\hat{U}_{(1)} + \hat{U}_{(L)}}{2} - \frac{a+b}{2} \right)^2 \right] \leq 2K^2 E \left[\left(\frac{U_{(1)} + U_{(L)}}{2} - \frac{a+b}{2} \right)^2 \right] + \epsilon_3 \quad (68)$$

where $\epsilon_3 = \epsilon_3(\delta_0, n)$ can be made arbitrarily small by making n sufficiently large and δ_0 sufficiently small.

Proof: Let \mathcal{B} be the event $\{\tilde{U}_i \neq \hat{U}_i, \forall i\}$. By inequality (20), we have $\Pr\{\mathcal{B}\} \leq \lambda$. Now,

$$\begin{aligned}
& K^2 E \left[\left(\frac{\hat{U}_{(1)} + \hat{U}_{(L)}}{2} - \frac{a+b}{2} \right)^2 - 2 \left(\frac{U_{(1)} + U_{(L)}}{2} - \frac{a+b}{2} \right)^2 \right] \\
&= \frac{K^2}{4} E \left[\left(\hat{U}_{(1)} + \hat{U}_{(L)} - (a+b) \right)^2 - 2 \left(U_{(1)} + U_{(L)} - (a+b) \right)^2 \right] \\
&= \frac{K^2}{4} E \left[\left((\hat{U}_{(1)} + \hat{U}_{(L)}) - (U_{(1)} + U_{(L)}) - ((a+b) - (U_{(1)} + U_{(L)})) \right)^2 - 2 \left(U_{(1)} + U_{(L)} - (a+b) \right)^2 \right] \\
&\leq \frac{K^2}{4} E \left[2 \left((\hat{U}_{(1)} + \hat{U}_{(L)}) - (U_{(1)} + U_{(L)}) \right)^2 + 2 \left((a+b) - (U_{(1)} + U_{(L)}) \right)^2 - 2 \left(U_{(1)} + U_{(L)} - (a+b) \right)^2 \right] \\
&= \frac{K^2}{2} E \left[\left((\hat{U}_{(1)} + \hat{U}_{(L)}) - (U_{(1)} + U_{(L)}) \right)^2 \right] \\
&\leq K^2 E \left[\left((U_{(1)} + U_{(L)}) - (\tilde{U}_{(1)} + \tilde{U}_{(L)}) \right)^2 \right] + K^2 E \left[\left((\tilde{U}_{(1)} + \tilde{U}_{(L)}) - (\hat{U}_{(1)} + \hat{U}_{(L)}) \right)^2 \right] \\
&\leq 2K^2 E \left[\left(U_{(1)} - \tilde{U}_{(1)} \right)^2 \right] + 2K^2 E \left[\left(U_{(L)} - \tilde{U}_{(L)} \right)^2 \right] + 2K^2 E \left[\left(\tilde{U}_{(1)} - \hat{U}_{(1)} \right)^2 \right] + 2K^2 E \left[\left(\tilde{U}_{(L)} - \hat{U}_{(L)} \right)^2 \right],
\end{aligned}$$

where $\tilde{U}_{(i)}$ are the order statistics of \tilde{U}_i , and the last two inequalities follow from the fact that $E[(A+B)^2] \leq 2E[A^2] + 2E[B^2]$.

Now, choose δ_0 to be sufficiently small to ensure that the ordering of variates U_i is preserved under quantization. Then, $U_{(1)}$ and $\tilde{U}_{(1)}$ correspond to the same agent's data, say the ℓ th agent. Therefore,

$$E \left[\left(U_{(1)} - \tilde{U}_{(1)} \right)^2 \right] = E \left[\left(U_\ell - \tilde{U}_\ell \right)^2 \right] \leq \delta_0$$

by (15). Similarly, $E \left[\left(U_{(L)} - \tilde{U}_{(L)} \right)^2 \right] \leq \delta_0$. Also, define $\tilde{u}_{\max} = \max\{|\tilde{u}| : \tilde{u} \in \tilde{\mathcal{U}}\}$. Now, for $i = \{1, \dots, L\}$:

$$\begin{aligned}
E \left[\left(\tilde{U}_{(i)} - \hat{U}_{(i)} \right)^2 \right] &= \sum_{u, u'} \left(\tilde{U}_{(i)} - \hat{U}_{(i)} \right)^2 \Pr \left\{ \tilde{U}_{(i)} = u, \hat{U}_{(i)} = u' \right\} \\
&= \sum_{u, u'} \left(\tilde{U}_{(i)} - \hat{U}_{(i)} \right)^2 \Pr \left\{ \tilde{U}_{(i)} = u, \hat{U}_{(i)} = u' \right\} \\
&\leq \sum_{u, u'} 4\tilde{u}_{\max}^2 \Pr \left\{ \tilde{U}_{(i)} = u, \hat{U}_{(i)} = u' \right\} \\
&= 4\tilde{u}_{\max}^2 \Pr \left\{ \tilde{U}_{(i)} \neq \hat{U}_{(i)} \right\} \\
&\leq 4\tilde{u}_{\max}^2 \Pr \left\{ \mathcal{B} \right\} \\
&\leq 4\tilde{u}_{\max}^2 \lambda.
\end{aligned}$$

Therefore,

$$\begin{aligned}
K^2 E \left[\left(\frac{\hat{U}_{(1)} + \hat{U}_{(L)}}{2} - \frac{a+b}{2} \right)^2 - 2 \left(\frac{U_{(1)} + U_{(L)}}{2} - \frac{a+b}{2} \right)^2 \right] &\leq 4K^2 \delta_0 + 8K^2 \tilde{u}_{\max}^2 \Pr(\mathcal{B}) \\
&\leq 4K^2 \delta_0 + 8K^2 \tilde{u}_{\max}^2 \lambda.
\end{aligned}$$

Now, choosing a sufficiently large n such that

$$\lambda < \frac{\epsilon_3 - 4K^2 \delta_0}{8K^2 \tilde{u}_{\max}^2},$$

yields the desired result. ■

Proposition 5. *The following inequality:*

$$\frac{\sum_{i=1}^n p_i A_i}{\sum_{i=1}^n p_i B_i} \geq \min_i \left(\frac{A_i}{B_i} \right) \quad (69)$$

holds, if $p_i, A_i, B_i \geq 0$ and not all are 0.

Proof: Let $m = \min_i \left(\frac{A_i}{B_i} \right)$. By definition,

$$\begin{aligned} A_i &\geq B_i m, \text{ for all } i = 1, \dots, n \\ \implies p_i A_i &\geq p_i B_i m, \text{ for all } i = 1, \dots, n \\ \implies \sum_{i=1}^n p_i A_i &\geq m \sum_{i=1}^n p_i B_i \\ \implies \frac{\sum_{i=1}^n p_i A_i}{\sum_{i=1}^n p_i B_i} &\geq m = \min_i \left(\frac{A_i}{B_i} \right). \end{aligned}$$

■